Extraction of mth Roots in Matrix Rings over Fields

Daniel E. Otero
Department of Mathematics
Syracuse University
Syracuse, New York 13244

Submitted by Richard A. Brualdi

ABSTRACT

A matrix $A \in M_n(F)$, F an arbitrary field with characteristic p (not necessarily positive), is an mth power in $M_n(F)$ if and only if each of its $p(x)$-primary components is, where $p(x)$ runs through the irreducible factors of the minimal or characteristic polynomial of A. This paper establishes necessary and sufficient criteria for determining when such a $p(x)$-primary matrix is an mth power. The criteria fall into three cases: (1) $p(x) = x$: If $e_1 \geq e_2 \geq \cdots$ is the sequence of exponents of $p(x)$ which form the elementary divisors of A, extended by adding infinitely many 0 terms, the criterion states that for all $i \geq 1$, $e_{(i-1)m+1} - e_{im} = 0$ or 1. (2) $p(x) \neq x$, m not divisible by p: If β is a root of the separable core $q(x)$ of $p(x)$ and $E = F(\beta)$, then the multiplicity of each elementary divisor of A must be representable as a sum of integers, not necessarily distinct, each of which is the degree of some irreducible factor over E of the polynomial $x^m - \beta$. (3) $p(x) \neq x$, m a power of p: In this case, the criterion combines the criteria in the first two cases: If $e_1 \geq e_2 \geq \cdots$ is the sequence of exponents of $p(x)$ which form the elementary divisors of A, extended by adding infinitely many 0 terms, then for all $i \geq 1$, $e_{(i-1)m+1} - e_{im} = 0$ or 1 and the multiplicity of each e_i must be representable as a sum of integers, not necessarily distinct, each of which is the degree of some irreducible factor over E of the polynomial $x^m - \beta$.

INTRODUCTION

Fix some arbitrary field F, and let $M_n(F)$ be the ring of $n \times n$ matrices over F. We concern ourselves in this paper with the problem of establishing criteria to determine when the matrix equation $X^m = A$, $A \in M_n(F)$, X an indeterminate over $M_n(F)$, has a solution.

655 Avenue of the Americas, New York, NY 10010 0024-3795/90/$3.50

This problem is one in a long line of investigations into solutions of matrix equations. However, it seems that only a few special cases of this problem have appeared in the literature, the best-known being the treatments of Wedderburn [1, pp. 119–122] and Gantmacher [2, pp. 232–239], both working over $F = \mathbf{C}$ (see also [3, p. 193] and [4]). Cross and Lancaster [5] solve the problem over $\mathbf{C}$ when $m = 2$ with criteria given in terms of a sequence called the *ascent sequence* of a matrix. Hodges [6] considers the special case $F = \mathrm{GF}(q)$, $A = I$, the identity matrix. Others, like [7], specialize to $m = 2$ and are concerned with particular choices of A; in the case of [8], the authors are only interested in solutions to $X^m = A$ that possess the same special properties as A. In this paper, we are able to give a criterion to determine the existence of solutions to $X^m = A$ for a given A with m and F arbitrary.

NOTATION.

(1) All polynomials we consider are monic unless otherwise indicated.
(2) We write $M^{(n)}$ to denote the direct sum of n copies of the module M.

1. THE MODULE-THEORETIC FORMULATION

A number of simple observations are apparent from the outset. In particular, since $X^m = A$ has a solution $X = B$ if and only if $X^m = UAU^{-1}$ has a solution $X = UBU^{-1}$ for all invertible $U \in M_n(F)$, the criteria we seek will not depend solely on A, but on similarity invariants of A. Further, the desired criteria will also have to depend intrinsically on F. For instance,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is easily seen not to be a square in $M_2(F)$ whenever F is a real subfield of the complex numbers $\mathbf{C}$; but over $\mathbf{C}$, it is the square of

$$\begin{pmatrix} \dfrac{1+i}{2} & \dfrac{1-i}{2} \\ \dfrac{1-i}{2} & \dfrac{1+i}{2} \end{pmatrix}.$$

As the elementary divisors of the matrix A form a collection of similarity

invariants that depend on F as well, one might expect our results to be expressed in terms of this data for A. This is precisely the case.

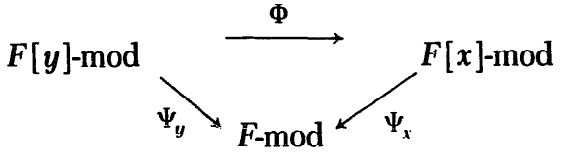
Moreover, we will benefit (by way of shortening proofs and clarifying ideas) by reformulating our matrix problem into an equivalent one expressed in module-theoretic language. If x is an indeterminate over F, the equivalence we speak of associates to the similarity class $\mathrm{sim}(A)$ of the matrix $A \in M_n(F)$ the (isomorphism class of) the $F[x]$-module $M = \mathrm{mod}_x(A)$, obtained by defining on an n-dimensional F-vector space V the action of x by the F-linear transformation on V whose matrix representation (in some basis) is A. This module is a finitely generated torsion module, annihilated by the ideal generated by the minimal polynomial of A. Indeed, the fundamental structure theorem for finitely generated torsion modules over a PID says that M is (isomorphic to) the direct sum of quotients of $F[x]$ by powers of prime ideals; we call this the *elementary-divisor decomposition* of M. These prime-power ideals, being principal, are the elementary divisors of M, or of A. Their product is the characteristic polynomial of A, whose degree is n. (The reader can consult any recent linear-algebra text for the details of this development, e.g., [9, Chapter 3].)

Matrices in $M_n(F)$ are similar if and only if they have the same set of elementary divisors, so the association $\mathrm{sim}(A) \to M$ is well defined and one-to-one. On the other hand, given an arbitrary finitely generated torsion $F[x]$-module M, the direct summands $F[x]/p(x)^eF[x]$ of its elementary divisor decomposition are vector spaces over F of dimension $\deg p(x)^e$, whence $\dim_F M$ is the degree n of the polynomial which is the product of the elementary divisors $p(x)^e$ (the characteristic polynomial). If V is the underlying n-dimensional vector space of M, the action of x defines an endomorphism of V whose matrix representations A run through a similarity class in $M_n(F)$. It is clear that $M = \mathrm{mod}_x(A)$, so $\mathrm{sim}(A) \to \mathrm{mod}_x(A)$ is bijective.

The reformulation of our original problem in terms of modules proceeds as follows. Let $X = B$ be a solution to $X^m = A$, $A \in M_n(F)$, and put $N = \mathrm{mod}_y(B)$, where y is a new indeterminate over F. Then the structure of M as an $F[x]$-module is identical to the structure of N as an $F[y^m]$-module. In other words, where $F[y]$-mod is the category of finitely generated torsion modules over $F[y]$, the F-algebra homomorphism $f: F[x] \to F[y]$ determined by $f(x) = y^m$ induces a functor $\Phi_m: F[y]\text{-mod} \to F[x]\text{-mod}$ that carries N to M. *Our problem therefore is to determine the image of the functor* $\Phi = \Phi_m$.

Let Ψ_x be the forgetful functor $F[x]\text{-mod} \to F\text{-mod}$ induced by the natural injection of rings $F \to F[x]$; $\Psi_x(M)$ is the underlying vector space of M. Similarly, $\Psi_y(N)$ is the underlying vector space of N. Since both vector spaces are of the same dimension over F (the actions of the indeterminates

on these vector spaces are endomorphisms with $n \times n$ matrix representations), they are isomorphic, so we freely identify them. Under this interpretation, the diagram of functors



commutes.

One important property of Φ we will use repeatedly is its additivity: $\Phi(N_1 \oplus N_2) = \Phi(N_1) \oplus \Phi(N_2)$ for all $N_1, N_2 \in F[y]$-mod. Indeed, more generally, Φ is exact: if $\varphi: N_1 \to N_2$ is an $F[y]$-module map which is injective/surjective, then $\Phi(\varphi): \Phi(N_1) \to \Phi(N_2)$ is an $F[x]$-module map which is injective/surjective. This is easily seen from the fact that $\Psi_x \Phi(\varphi) = \Psi_y(\varphi)$.

2. REDUCTION TO THE PRIMARY CASE

In order to develop the criteria that will determine whether a given $M \in F[x]$-mod is the image under $\Phi = \Phi_m$ of some $N \in F[y]$-mod, we approach by way of a number of reductions.

Suppose that

$$\begin{aligned} &p_1(x)^{e_{11}}, p_1(x)^{e_{12}}, \ldots, p_1(x)^{e_{1s_1}}, \\ &p_2(x)^{e_{21}}, p_2(x)^{e_{22}}, \ldots, p_2(x)^{e_{2s_2}}, \\ &\vdots \\ &p_t(x)^{e_{t1}}, p_t(x)^{e_{t2}}, \ldots, p_t(x)^{e_{ts_t}} \end{aligned}$$

is the list of elementary divisor polynomials of M, where $p_1(x), p_2(x), \ldots, p_t(x)$ are the distinct irreducible factors over F of the minimal polynomial of M (which generates $\operatorname{ann}_{F[x]} M$). For $i = 1, 2, \ldots, t$, let M_i be the submodule of M which is the direct sum of those indecomposable submodules of M (the distinct summands in the elementary divisor decomposition) whose annihilators are powers of the ideal generated by $p_i(x)$. We call M_i the *$p_i(x)$-primary component* of M, and $M_1 \oplus \cdots \oplus M_t$ its *primary decomposition*. When $t = 1$ and $p(x) = p_1(x)$, we say that M is a *$p(x)$-primary* module, or simply that M is primary; this is equivalent to saying that the annihilator of M is the power of an irreducible polynomial.

THEOREM 1. *M lies in the image of Φ if and only if each of its primary components lies in the image of Φ.*

Proof. We need only prove the necessity; the sufficiency is clear from the additivity of Φ.

Let $M = \Phi(N)$, and suppose $M = M_1 \oplus \cdots \oplus M_t$ is its primary decomposition. Then $V_i = \Psi_x(M_i)$ is a vector subspace of $V = \Psi_y(N) = \Psi_x(M)$. It suffices to show that the linear transformation of V defined by multiplication by y leaves V_i invariant, for this will prove the existence of an $F[y]$-submodule N_i of N for which $\Psi_y(N_i) = \Psi_x(M_i)$. Since $y^m = x$ on V, it will also follow that $\Phi(N_i) = M_i$, completing the argument.

Put $L_i = M_1 \oplus \cdots \oplus M_{i-1} \oplus M_{i+1} \oplus \cdots \oplus M_t$. Then $\operatorname{ann}_{F[x]} L_i$ and $\operatorname{ann}_{F[x]} M_i$ are coprime ideals in $F[x]$. If $q_i(x)$ generates $\operatorname{ann}_{F[x]} L_i$ and $p_i(x)^{e_i}$ generates $\operatorname{ann}_{F[x]} M_i$, then there exist polynomials $a(x)$ and $b(x)$ for which $a(x)q_i(x) + b(x)p_i(x)^{e_i} = 1$.

Pick $u \in M_i$. Then there exist unique $v \in L_i$ and $w \in M_i$ for which $yu = v + w$. Since $xy = y^{m+1} = yx$,

$$\begin{aligned} yu &= y\left[a(x)q_i(x) + b(x)p_i(x)^{e_i}\right]u \\ &= ya(x)q_i(x)u \\ &= a(x)q_i(x)yu \\ &= a(x)q_i(x)[v + w] \\ &= a(x)q_i(x)w \\ &\in M_i, \end{aligned}$$

so we are done. ∎

We may assume then that M is a $p(x)$-primary module with $p(x)$ an irreducible over F. Our next reduction will show that we can assume that $p(x)$ is purely inseparable over F. This is accomplished by another functorial construction.

3. THE FUNCTOR Ω

For every irreducible $p(x)$ over F there is an associated polynomial $q(x)$, separable over F, which we call the *separable core* of $p(x)$, or simply the

core of $p(x)$, defined by the condition $p(x) = q(x^\rho)$, where $\rho = p^r$ is a power of the characteristic of F [10, p. 179]. A number of observations are apparent. First, the core of $p(x)$ is equal to $p(x)$ unless F is an imperfect field. Also, it is necessarily irreducible. Next, the degree of $q(x)$ is the *separable degree* of $p(x)$, and ρ is the *inseparable degree* of $p(x)$ [so that $\deg p(x) = \rho \deg q(x)$]. Therefore, the roots of $p(x)$ (lying in some fixed algebraic closure of F) each occur with multiplicity ρ, and their pth powers are precisely the roots of $q(x)$, which are distinct.

Let α be a root of $p(x)$, so that $\beta = \alpha^\rho$ is a root of its core $q(x)$, and let E denote the extension field $F(\beta)$. If X is an indeterminate over E, the polynomial $X^\rho - \beta \in E[X]$ is irreducible, for it factors over the algebraic closure as $(X-\alpha)^\rho$ and $\alpha \in E$ if and only if $\rho = 1$. We now construct a functor Ω between the full subcategory $(F[x], p(x))$-mod of $F[x]$-mod, whose objects are, by definition, the $p(x)$-primary modules with elementary divisor decompositions of the form

$$F[x]/p(x)^{e_1}F[x] \oplus F[x]/p(x)^{e_2}F[x] \oplus \cdots \oplus F[x]/p(x)^{e_t}F[x], \quad (1)$$

and the full subcategory $(E[X], X^\rho - \beta)$-mod of $E[X]$-mod, whose objects are the $(X^\rho - \beta)$-primary modules which have the similar form

$$E[X]/(X^\rho - \beta)^{e_1}E[X] \oplus E[X]/(X^\rho - \beta)^{e_2}E[X]$$
$$\oplus \cdots \oplus E[X]/(X^\rho - \beta)^{e_t}E[X]. \quad (2)$$

Because of the decomposition theorem, we can define Ω on modules (the objects of the category) by the rule that maps the $F[x]$-module (1) to the $E[X]$-module (2). This implies that Ω is additive and bijective on objects.

To define Ω on morphisms, we first note that if

$$M_1 = F[x]/p(x)^{d_1}F[x] \oplus F[x]/p(x)^{d_2}F[x] \oplus \cdots \oplus F[x]/p(x)^{d_s}F[x]$$

and

$$M_2 = F[x]/p(x)^{e_1}F[x] \oplus F[x]/p(x)^{e_2}F[x] \oplus \cdots \oplus F[x]/p(x)^{e_t}F[x],$$

then

$$\operatorname{Hom}_{F[x]}(M_1, M_2) = \bigoplus \operatorname{Hom}_{F[x]}\left(F[x]/p(x)^{d_i}F[x], F[x]/p(x)^{e_j}F[x]\right),$$

so that any morphism $\varphi: M_1 \to M_2$ can be represented by an $s \times t$ matrix whose (i, j) entry is a morphism

$$\varphi_{ij}: F[x]/p(x)^{d_i}F[x] \to F[x]/p(x)^{e_j}F[x].$$

Composition of morphisms is then multiplication of matrices. Since $\Omega(\varphi): \Omega(M_1) \to \Omega(M_2)$ will have a similar representation as an $s \times t$ matrix whose (i, j) entry is an element of

$$\mathrm{Hom}_{E[X]}\big(E[X]/(X^\rho - \beta)^{d_i}E[X], E[X]/(X^\rho - \beta)^{e_j}E[X]\big),$$

we will simply define $\Omega(\varphi)$ for morphisms φ between indecomposables in $(F[x], p(x))$-mod and require that $(\Omega(\varphi))_{ij} = \Omega(\varphi_{ij})$.

Suppose then that $\varphi \in \mathrm{Hom}_{F[x]}(F[x]/p(x)^dF[x], F[x]/p(x)^eF[x])$. Since $F[x]/p(x)^dF[x]$ is a cyclic module with generator $1 + p(x)^dF[x]$, we find that φ is uniquely determined by $\varphi(1 + p(x)^dF[x])$. Write

$$\varphi\big(1 + p(x)^dF[x]\big) = f(x) + p(x)^eF[x]$$

for some $f(x) \in F[x]$. As each element of

$$\mathrm{Hom}_{E[X]}\big(E[X]/(X^\rho - \beta)^dE[X], E[X]/(X^\rho - \beta)^eE[X]\big)$$

is similarly uniquely determined by its action on the element $(1 + (X^\rho - \beta)^dE[X])$, we define $\Omega(\varphi)$ by putting

$$\Omega(\varphi)\big(1 + (X^\rho - \beta)^dE[X]\big) = f(X) + (X^\rho - \beta)^eE[X].$$

If $\varphi(1 + p(x)^dF[x]) = g(x) + p(x)^eF[x]$ for some other representative $g(x) \in F[x]$, then $p(x)^e$ would divide the difference $f(x) - g(x)$; but $p(x)$ factors over E as a multiple of $X^\rho - \beta$, so $f(X) - g(X)$ is a multiple of $(X^\rho - \beta)^e$. Hence $\Omega(\varphi)$ is well defined.

Observe also that if φ' is another element of $\mathrm{Hom}_{F[x]}(F[x]/p(x)^dF[x], F[x]/p(x)^eF[x])$, we have $\Omega(\varphi + \varphi') = \Omega(\varphi) + \Omega(\varphi')$, and if $\psi \in \mathrm{Hom}_{F[x]}(F[x]/p(x)^cF[x], F[x]/p(x)^dF[x])$ is determined by

$$\psi\big(1 + p(x)^cF[x]\big) = g(x) + p(x)^dF[x],$$

$g(x) \in F[x]$, then

$$\begin{aligned}\Omega(\varphi)\Omega(\psi)\big(1+(X^\rho-\beta)^c F[X]\big) &= g(X)\Omega(\varphi)\big(1+(X^\rho-\beta)^d F[X]\big)\\ &= f(X)g(X)+(X^\rho-\beta)^e F[X]\\ &= \Omega(\varphi\psi)\big(1+(X^\rho-\beta)^c F[X]\big),\end{aligned}$$

whence $\Omega(\varphi\psi)=\Omega(\varphi)\Omega(\psi)$.

The verification that $\Omega(\mathrm{id}_M)=\mathrm{id}_{\Omega(M)}$ for all $M \in (F[x], p(x))$-mod is immediate; and, given $M_1, M_2, M_3 \in (F[x], p(x))$-mod and morphisms $\psi: M_1 \to M_2$ and $\varphi: M_2 \to M_3$, the fact that $\Omega(\varphi\psi)=\Omega(\varphi)\Omega(\psi)$ is just matrix multiplication. So Ω is a well-defined functor from $(F[x], p(x))$-mod to $(E[X], X^\rho-\beta)$-mod.

We have already mentioned that Ω is bijective on objects; we will show that it is also bijective on morphisms, i.e., that Ω is an isomorphism of categories. We note that in the light of the remarks above, it suffices to prove that Ω is a bijection between $\mathrm{Hom}_{F[x]}(F[x]/p(x)^dF[x], F[x]/p(x)^eF[x])$ and $\mathrm{Hom}_{E[X]}(E[X]/(X^\rho-\beta)^dE[X], E[X]/(X^\rho-\beta)^eE[X])$. To do this, we construct a map $\omega: F[x]/p(x)^eF[x] \to E[X]/(X^\rho-\beta)^eE[X]$ by

$$\omega\big(f(x)+p(x)^eF[x]\big)=f(X)+(X^\rho-\beta)^eE[X].$$

Now ω is a homomorphism of rings; indeed, $\omega(a+p(x)^eF[x])=a+(X^\rho-\beta)^eE[X]$ for all $a \in F$, so that ω is an F-algebra homomorphism. Since $f(x)+p(x)^eF[x] \in \ker\omega$ implies that $f(X)$ is divisible by $(X^\rho-\beta)^e$ over E so that $f(x)$ has α as a root, then $f(x)$ is divisible by $p(x)^e$ over F, and we conclude that ω is injective. But

$$\begin{aligned}\dim_F F[x]/p(x)^eF[x] &= e[F(\alpha):F] = e[F(\alpha):E][E:F] = e\rho[E:F]\\ &= [E:F]\dim_E E[X]/(X^\rho-\beta)^eE[X]\\ &= \dim_F E[X]/(X^\rho-\beta)^eE[X],\end{aligned}$$

so ω is surjective as well. So, if $\varphi \in \mathrm{Hom}_{F[x]}(F[x]/p(x)^dF[x], F[x]/p(x)^eF[x])$ is determined by

$$\varphi\big(1+p(x)^dF[x]\big)=f(x)+p(x)^eF[x],$$

$f(x) \in F[x]$, it follows that $\Omega(\varphi)$ is determined by

$$\Omega(\varphi)\big(1+(X^p-\beta)^d E[X]\big) = \omega\big(f(x)+F[x]/p(x)^e F[x]\big).$$

It is clear that since ω is an isomorphism, Ω is bijective on morphisms.

Ostensibly, it seems that the definition of Ω is dependent on the choice of the root β of $q(x)$, but in fact if β' is another such root, $E' = F(\beta')$, and $\Omega' : (F[x], p(x))\text{-mod} \to (E'[X], X^p - \beta')\text{-mod}$ is the associated functor, then because the categories $(E[X], X^p-\beta)$-mod and $(E'[X], X^p-\beta')$-mod are naturally isomorphic, the isomorphism being induced by the isomorphism of fields $E \to E'$, we have a commutative diagram

$$\begin{array}{ccc} & (F[x], p(x))\text{-mod} & \\ {\scriptstyle\Omega}\swarrow & & \searrow{\scriptstyle\Omega'} \\ (E[X], X^p-\beta)\text{-mod} & \cong & (E'[X], X^p-\beta')\text{-mod} \end{array}$$

of categories. In this sense, then, the choice of the root of $q(x)$ used to define Ω is immaterial; the definition of Ω depends only on the polynomial $p(x)$.

THEOREM 2. *If $M \in (F[x], p(x))$-mod is the image under the functor $\Phi : F[y]\text{-mod} \to F[x]\text{-mod}$ of $N \in F[y]$-mod, then $\Omega(M) \in (E[X], X^p-\beta)$-mod is the image under the functor $\Phi : E[Y]\text{-mod} \to E[X]\text{-mod}$ of some $N' \in E[Y]$-mod, and conversely.*

Proof. If $M = \Phi(N)$, then the transformations on $V = \Psi_x(M) = \Psi_y(N)$ given by the actions of y and x commute with each other, so $y \in \operatorname{End}_{F[x]} M$. Since M has the form (1), we can represent y as a $t \times t$ matrix (y_{ij}), where

$$y_{ij} \in \operatorname{Hom}_{F[x]}\big(F[x]/p(x)^{e_i}F[x], F[x]/p(x)^{e_j}F[x]\big);$$

note also that $(y_{ij})^m$ is a $t \times t$ matrix representation for x.

The element $\Omega(y) = (\Omega(y_{ij})) \in \operatorname{End}_{E[X]} \Omega(M)$ defines an $E[X]$-module map whose mth power is X, so the $E[Y]$-module structure on $\Omega(M)$ defined by having Y act like $\Omega(y)$ is an element of $E[Y]$-mod whose image under $\Phi : E[Y]\text{-mod} \to E[X]\text{-mod}$ is $\Omega(M)$. This proves the first statement of the theorem. The converse follows from the fact that since Ω is an isomorphism of categories, this argument is reversible. ∎

We may therefore assume, for the sake of our analysis, that the $p(x)$-primary module M is such that $p(x)$ is purely inseparable over F, that is, $p(x) = x^p - b$, $b \in F$. When F is a perfect field, this condition is equivalent

to saying that $\deg p(x) = 1$. In general, however, as we show in the next section, we can still reduce to the degree-one case.

4. REDUCTION TO THE CASE $p(x) = x - a$

In what follows, we reserve the lowercase Greek letters ρ, σ, τ to stand for the integers p^r, p^s, p^t, with $r, s, t \geqslant 0$, respectively.

Our goal in this section is to show that we may further reduce our problem from the consideration of a purely inseparable $p(x)$ to the simpler case $p(x) = x - a$. As a result, *we assume throughout this section that F is imperfect*, hence that $\operatorname{char} F = p > 0$. Proposition 3 below sets aside a useful calculation of the "ρth power" of a module. Then, in Theorem 4, we show how to obtain the reduction. This reduction is functorial, of course, so we require the definition of a new functor Γ_ρ, different from the "mth power" functor Φ_m and the forgetful functors $\Psi_*: F[*]\text{-mod} \to F\text{-mod}$ we have been using. The proof of Theorem 4 boils down to combining the result of Proposition 3—which says basically that $\Phi_\rho \Gamma_\rho(M) = M^{(\rho)}$ and that Φ_ρ is bijective on objects—with the commutativity relation $\Phi_m \Phi_\rho = \Phi_\rho \Phi_m$.

PROPOSITION 3. *Let $\Phi_\rho: F[Z]\text{-mod} \to F[z]\text{-mod}$ be the "ρth power" functor, i.e., $\Phi_\rho(N)$ is the $F[z]$-module obtained by having z act as Z^ρ on $N \in F[Z]$-mod. If $p(Z)$ is irreducible over F with core $q(Z)$ and inseparable degree $\sigma \geqslant \rho$, then*

$$\Phi_\rho\big(F[Z]/p(Z)^e F[Z]\big) = \big(F[z]/q(z^{\sigma/\rho})^e F[z]\big)^{(\rho)}. \tag{3}$$

Conversely, if $P \in F[Z]$-mod satisfies $\Phi_\rho(P) = (F[z]/q(z^{\sigma/\rho})^e F[z])^{(\rho)}$, then $P = F[Z]/p(Z)^e F[Z]$.

Proof. To prove the first statement, put $P = F[Z]/p(Z)^e F[Z] \in F[Z]$-mod, $p(Z)$ an irreducible polynomial with inseparable degree $\sigma = p^s$ and core $q(Z)$ of degree d. Viewed as a linear transformation of the vector space $V = \Psi_Z(P)$, which has basis $B = \{1, Z, Z^2, \ldots, Z^{\sigma de - 1}\}$ (we write Z^i for $Z^i + p(Z)^e F[Z]$), $z = Z^\rho$ leaves invariant the subspace V_i with basis $B_i = \{Z^j \mid j \equiv i \pmod{\rho}\}$, $i = 0, 1, \ldots, \rho - 1$. Since B is the disjoint union of the B_i, $V = \oplus V_i$; in addition, there exists a submodule Q_i of $\Phi_\rho(P)$ with $V_i = \Psi_z(Q_i)$ that satisfies $\Phi_\rho(P) = \oplus Q_i$. Moreover, Q_i is a cyclic $F[z]$-module with generator Z^i.

Since $p(Z)^e$ annihilates V, $q(z^{\sigma/\rho})^e$ annihilates each V_i, so the restriction of the endomorphism z to V_i has minimal polynomial dividing $q(z^{\sigma/\rho})^e$. Now

$q(z^{\sigma/\rho})$ is irreducible in $F[z]$, else we can replace the variable z with z^{ρ} to contradict the irreducibility of $q([z^{\rho}]^{\sigma/\rho}) = p(z)$. So the endomorphism z has as minimal polynomial over V_i a power of $q(z^{\sigma/\rho})$, say the e_i power, $e_i \leqslant e$. Consequently,

$$Q_i = F[z]/q(z^{\sigma/\rho})^{e_i}F[z].$$

However, if any one of the $e_i < e$, we have

$$\sigma de = \operatorname{card} B = \dim_F V = \sum \dim_F V_i = \sum \frac{\sigma}{\rho} de_i < \frac{\sigma}{\rho} d\rho e = \sigma de,$$

so in fact each e_i equals e. So $Q_i = F[z]/q(z^{\sigma/\rho})^{e}F[z]$ for each i, which proves the first half of the proposition.

Now suppose that $P \in F[Z]$-mod satisfies

$$\Phi_{\rho}(P) = \left(F[z]/q(z^{\sigma/\rho})^{e}F[z]\right)^{(\rho)}.$$

Then as an element of $\operatorname{End}_{F[Z]}P$, $p(Z)^{e} = q(z^{\sigma/\rho})^{e} = 0$, since $q(z^{\sigma/\rho})^{e}$ annihilates $V = \Psi_z(\{F[z]/q(z^{\sigma/\rho})^{e}F[z]\}^{(\rho)}) = \Psi_Z(P)$, so the minimal polynomial of Z is a power of the irreducible $p(Z)$. It follows that P is $p(Z)$-primary. By (3) and the additivity of Φ_{ρ}, we must have $P = F[Z]/p(Z)^{e}F[Z]$. ∎

The reduction we seek is effected by a functor $\Gamma_{\rho}: (F[x], x-b)$-mod $\to$ $(F[X], X^{\rho}-b)$-mod which we now define. Suppose the polynomial $x^{\rho}-b \in F[x]$ is irreducible, i.e., b is not a pth power in F. Then the action of Γ_{ρ} on objects is given by setting

$$\Gamma_{\rho}\left(F[x]/(x-b)^{e}F[x]\right) = F[X]/(X^{\rho}-b)^{e}F[X],$$

then extending additively via the elementary divisor decomposition (as we did for Ω in Section 3); for morphisms, it suffices to define $\Gamma_{\rho}(\varphi)$ for $\varphi \in \operatorname{Hom}(F[x]/(x-b)^{d}F[x], F[x]/(x-b)^{e}F[x])$, then extend by additivity for more general morphisms. Since

$$\varphi \in \operatorname{Hom}\left(F[x]/(x-b)^{d}F[x], F[x]/(x-b)^{e}F[x]\right)$$

is determined by

$$\varphi\left(1+(x-b)^{d}F[x]\right) = f(x)+(x-b)^{e}F[x],$$

we define $\Gamma_{\rho}(\varphi)$ by setting

$$\Gamma_{\rho}(\varphi)\left(1+(X^{\rho}-b)^{d}F[X]\right) = f(X^{\rho}) + (X^{\rho}-b)^{e}F[X].$$

We leave to the reader the verifications that Γ_ρ is a (well-defined) faithful functor which is bijective on objects.

THEOREM 4. *The module* $M \in (F[x], x-b)$-mod *lies in the image of the functor* $\Phi_m: F[y]\text{-mod} \to F[x]\text{-mod}$ *if and only if the module* $\Gamma_\rho(M) \in (F[X], X^\rho - b)$-mod *lies in the image of the functor* $\Phi_m: F[Y]\text{-mod} \to F[X]\text{-mod}$.

Proof. Suppose that $\Phi_m(N) = M$, and that $N = F[y]/p(y)^e F[y]$, where $p(y)$ is irreducible with core $q(y)$ and inseparable degree σ (not necessarily $\geqslant \rho$). Let $p_1(Y)$ be a nontrivial irreducible factor of the polynomial $q(Y^{\rho\sigma}) \in F[Y]$, and suppose that $p_1(Y)$ has core $q_1(Y)$ and inseparable degree τ. If α is a root of $p_1(Y)$, it is also a root of $q(Y^{\rho\sigma})$; in particular, since both the $\rho\sigma$th power and the τth power of α are separable elements over F, it must be that $\rho\sigma = \tau$. Therefore, $q_1(Y)$ divides $q(Y)$, and the irreducibility of $q(Y)$ implies $q_1(Y) = q(Y)$. This proves that $q(Y^{\rho\sigma})$ is irreducible over F. With $N' = F[Y]/q(Y^{\rho\sigma})^e F[Y]$, it follows from (3) that $\Phi_\rho(N') = N^{(\rho)}$.

If N has more than one summand in its elementary-divisor decomposition, we can use this argument with the additivity of Φ_ρ to find an $N' \in F[Y]$-mod, each of whose elementary-divisor polynomials is inseparable over F with degree of inseparability $\geqslant \rho$, for which $\Phi_\rho(N') = N^{(\rho)}$.

Now the diagram of functors

$$\begin{array}{ccc} F[y]\text{-mod} & \xrightarrow{\Phi_m} & F[x]\text{-mod} \\ \Big\downarrow{\scriptstyle \Phi_\rho} & & \Big\downarrow{\scriptstyle \Phi_\rho} \\ F[Y]\text{-mod} & \xrightarrow[\Phi_m]{} & F[X]\text{-mod} \end{array}$$

commutes, since the composition in either direction around the square is $\Phi_{m\rho}$. So by Proposition 3, $\Phi_\rho\Phi_m(N') = \Phi_m\Phi_\rho(N') = \Phi_m(N)^{(\rho)} = M^{(\rho)} = \Phi_\rho\Gamma_\rho(M)$, whence $\Phi_m(N') = \Gamma_\rho(M)$.

To prove the converse, suppose $\Gamma_\rho(M) = \Phi_m(N')$ for some $N' \in F[Y]$-mod with only one summand in its elementary divisor decomposition; say $N' = F[Y]/p(Y)^e F[Y]$. As an element of $\mathrm{End}_{F[Y]} N'$, multiplication by $(Y^{m\rho} - b)^k = (X^\rho - b)^k$ is the zero map when $k \geqslant e$, so $Y^{m\rho} - b$ is divisible by $p(Y)$. Rewrite $m\rho = n\sigma$, where n is prime to p. Then, if α is a root of $p(Y)$, it is also a root of $Y^{m\rho} - b = Y^{n\sigma} - b$, which has inseparable degree σ. Therefore, α occurs with multiplicity exactly σ as a root of $Y^{n\sigma} - b$, hence also as a root of $p(x)$, which is its minimal polynomial over F. It follows that

the inseparable degree of $p(Y)$ is $\sigma \geq \rho$, and if $q(Y)$ is its core, then $p(Y) = q(Y^{\sigma})$. Hence, by Proposition 3 and additivity of Φ_{ρ},

$$M^{(\rho)} = \Phi_{\rho}\Gamma_{\rho}(M) = \Phi_{\rho}\Phi_{m}(N')$$

$$= \Phi_{m}\Phi_{\rho}\big(F[Y]/p(Y)^{e}F[Y]\big)$$

$$= \Phi_{m}\Big(\big\{F[y]/q(y^{\sigma/\rho})^{e}F[y]\big\}^{(\rho)}\Big) \qquad \text{[again by (3)]}$$

$$= \Phi_{m}\big(F[y]/q(y^{\sigma/\rho})^{e}F[y]\big)^{(\rho)},$$

and on comparing the elementary-divisor decompositions of $M^{(\rho)}$ and $\Phi_{m}(F[Y]/q(Y^{\sigma/\rho})^{e}F[Y])^{(\rho)}$ in $F[X]$-mod, we conclude that $M = \Phi_{m}(F[Y]/q(Y^{\sigma/\rho})^{e}F[Y])$.

If N has more than one summand in its elementary-divisor decomposition, the general case follows immediately from the additivity of Φ_{m} and Φ_{ρ}. ∎

The results of Section 3 reduce our problem to the consideration of modules M from subcategories $(F[x], p(x))$-mod where $p(x)$ is irreducible and purely inseparable: $p(x) = x^{\rho} - b$ for some $b \in F$ which is not a pth power. Applying Theorem 4 and the bijectivity of Γ_{ρ} on objects, we can now reduce our analysis to the consideration of $M' = \Gamma_{\rho}(M) \in (F[x], x - b)$-mod, that is, we can take $p(x)$ to be of degree one.

5. REDUCTION TO THE SCALAR CASE

We can now take the elementary-divisor decomposition of M to be of the form

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_t, \qquad M_i = \big\{F[x]/(x-a)^{e_i}F[x]\big\}^{(s_i)},$$

$$\text{and} \quad e_1 > e_2 > \cdots > e_t. \tag{4}$$

Each of the submodules M_i of M in (4) is called a *homogeneous component* of M. Our next reduction (Theorem 7) will show that under certain conditions we may further assume that M has but one homogeneous component.

LEMMA 5. *Let a_k be the coefficient of X^k in the binomial series expansion*

$$(1+X)^{1/m} = 1 + a_1X + a_2X^2 + \cdots. \tag{5}$$

If p does not divide m, then each a_k lies in F.

Proof. Put $a_0 = 1$. Raising both sides of (5) to the mth power and comparing coefficients of X^k yields $1 = ma_1$, and for $k \geqslant 2$,

$$0 = ma_k + \sum a_{i_1}a_{i_2}\cdots a_{i_m},$$

where the summation is over all indices $0 \leqslant i_1, i_2, \ldots, i_m \leqslant k$ that sum to k. By induction, then, all a_k are rational numbers which in lowest terms have denominators prime to p, completing the proof. ∎

PROPOSITION 6. *Suppose that $a \neq 0$ and that p does not divide m. If $\{F[x]/(x-a)F[x]\}^{(s)} \in F[x]$-mod is in the image of Φ_m, then so is $\{F[x]/(x-a)^eF[x]\}^{(s)}$ for every $e \geqslant 1$.*

Proof. Let $M = F[x]/(x-a)^eF[x]$. Then $\Psi_x(M^{(s)}) \cong \Psi_x(M)^{(s)}$ is naturally isomorphic to $\Psi_x(M) \otimes_F F^{(s)}$. Under this isomorphism, the action of x on the tensor-product space is determined by $x(u \otimes v) = xu \otimes v$ for $u \in \Psi_x(M)$ and $v \in F^{(s)}$; that is, x acts on the second tensor factor $F^{(s)}$ trivially. Now in view of the lemma, we can define an element $\xi \in \operatorname{End}_{F[x]} M$ which is multiplication by

$$\sum_{k=0}^{e-1} a_k\left[a^{-1}(x-a)\right]^k = \sum_{k=0}^{\infty} a_k\left[a^{-1}(x-a)\right]^k, \tag{6}$$

where the a_k are defined as in the lemma, and equality in (6) holds because $(x-a)^e = 0$ in M. Consequently, $\xi^m = 1 + a^{-1}(x-a) = a^{-1}x$. (This use of the binomial-series trick is essentially the same as in [1, p. 30] or [2, p. 232].)

By hypothesis there exists an $F[y]$-module N which satisfies $\Phi_m(N) = \{F[x]/(x-a)F[x]\}^{(s)}$. Since the action of y on the s-dimensional vector space $\Psi_y(N)$ is given by an endomorphism η whose mth power is scalar

multiplication by a, we can build an $F[y]$-module structure on $\Psi_x(M)\otimes_F F^{(s)}$ via $y(u\otimes v)=\xi u\otimes \eta v$, whence

$$y^m(u\otimes v)=\xi^m u\otimes \eta^m v=a^{-1}xu\otimes av=xu\otimes v=x(u\otimes v).$$

This defines an $F[y]$-module whose image under Φ_m is $\{F[x]/(x-a)^eF[x]\}^{(s)}$, completing the proof. ∎

THEOREM 7. *Suppose that $a\neq 0$ and that p does not divide m. Then M as in* (4) *lies in the image of $\Phi=\Phi_m$ if and only if each of its homogeneous components does.*

Proof. By additivity of Φ, we need only prove the necessity. Suppose then that $M=\Phi(N)$, and let $P=(x-a)M$. Then $P=P_1\oplus P_2\oplus\cdots\oplus P_t$, where $P_i=(x-a)M_i$. Also, for each $i=1,2,\ldots,t$, write L_i for the $f[x]$-submodule

$$P_1\oplus P_2\oplus\cdots\oplus P_i\oplus M_{i+1}\oplus M_{i+2}\oplus\cdots\oplus M_t$$

of M.

Choose $u_i\in\Psi_x(M_i)$ for some i; then, viewing y as a transformation on $\Psi_y(N)=\Psi_x(M)$, we have a unique expression $yu_i=v_1+v_2+\cdots+v_t$ with $v_j\in\Psi_x(M_j)$. Therefore,

$$(x-a)^{e_i}yu_i=(x-a)^{e_i}v_1+(x-a)^{e_i}v_2+\cdots+(x-a)^{e_i}v_{i-1} \qquad (7)$$

on the one hand, and

$$(x-a)^{e_i}yu_i=y(x-a)^{e_i}u_i=0$$

on the other, so that each term on the right side of (7) is 0. From this we conclude that $v_j\in\Psi_x(P_j)$ for all $j<i$, i.e., that $y\Psi_x(M_i)\subset\Psi_x(L_i)$.

Now if $u\in\Psi_x(P)$, then $u=(x-a)w$ for some $w\in\Psi_x(M)$, so $yu=y(x-a)w=(x-a)yw\in(x-a)\Psi_x(M)=\Psi_x(P)$, that is, $y\Psi_x(P)\subset\Psi_x(P)\subset\Psi_x(L_i)$. Therefore,

$$\begin{aligned}y\Psi_x(L_i)&=y\left[\Psi_x(P+M_i+M_{i+1}+\cdots+M_t)\right]\\&=y\left[\Psi_x(P)+\Psi_x(M_i)+\Psi_x(M_{i+1})+\cdots+\Psi_x(M_t)\right]\\&\subset\Psi_x(L_i).\end{aligned}$$

It follows that there is some $f[y]$-submodule N_i of N for which $L_i = \Phi(N_i)$.

We then have that for all $i \geq 1$, $\Phi(N_{i-1}/N_i) = L_{i-1}/L_i$, as Φ is exact, where we are taking $L_0 = M$ and $N_0 = N$. But

$$L_{i-1}/L_i = (L_i + M_i)/L_i$$

$$\cong M_i/(L_i \cap M_i) = M_i/P_i \cong \{F[x]/(x-a)F[x]\}^{(s_i)}$$

(as $F[x]$-modules), so we can apply Proposition 6 to conclude that M_i is in the image of Φ. ■

The reduction given by Theorem 7 requires the restrictions $a \neq 0$ and p prime to m. For instance, take F to be the real number field $\mathbf{R}$ and $a = 0$. Then the module

$$M = \mathbf{R}[x]/x\mathbf{R}[x] \oplus \mathbf{R}[x]/x^2\mathbf{R}[x]$$

is the image under Φ_2 of $N = \mathbf{R}[y]/y^3\mathbf{R}[y]$, which is equivalent in matrix language to saying that the similarity class (over $\mathbf{R}$) of matrices with Jordan canonical form

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

are squares of matrices in the similarity class determined by the single Jordan block

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

However, the homogeneous component $\mathbf{R}[x]/x^2\mathbf{R}[x]$ of M does not lie in the image of Φ_2, since the matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

is not a square.

Consequently, our analysis must take a different tack to handle those degree-one primary modules that do not satisfy $a \neq 0$ and p prime to m. Those for which $a = 0$ are considered in the next section, while those for

which $a \neq 0$ and p divides m are studied in Section 7. (Indeed, Proposition 3 is a part of the analysis of this last case.)

However, when $a \neq 0$ and p is prime to m, by Theorem 7 we can assume that the degree-one primary module M is itself *homogeneous*, i.e., is its own only homogeneous component. So M is as in (4) with $t = 1$:

$$M = \{F[x]/(x-a)^e F[x]\}^{(s)}.$$

The final reduction will show that we may assume that $e = 1$, i.e., that the action of x on $\Psi_x(M)$ is multiplication by the scalar a.

THEOREM 8. *Suppose that $a \neq 0$ and that p does not divide m. Then for all $e \geqslant 1$, the module $M = \{F[x]/(x-a)^e F[x]\}^{(s)}$ lies in the image of Φ if and only if the module $\{F[x]/(x-a)F[x]\}^{(s)}$ lies in the image of Φ.*

Proof. The sufficiency is precisely Proposition 6, so we need only prove the necessity. As in the proof of Theorem 7, if $P = (x-a)M$, then $y\Psi_x(P) \subset \Psi_x(P)$. So there is a submodule Q of N for which $\Psi_y(Q) = \Psi_x(P)$ and $Q = \Phi(P)$. Therefore, $\Phi(N/Q) = M/P \cong \{F[x]/(x-a)F[x]\}^{(s)}$. ∎

For the module $\{F[x]/(x-a)F[x]\}^{(s)}$—which we call a *scalar module of size s*, since the action of x on $\Psi_x(M)$ is multiplication by the scalar a—it is easy to give the criterion for lying in the image of Φ.

THEOREM 9. *Suppose that $a \neq 0$ and that p does not divide m. Then the scalar module $M = \{F[x]/(x-a)F[x]\}^{(s)}$ lies in the image of Φ if and only if s can be written as a sum of degrees of irreducible factors over F of $y^m - a$. (These factors need not be distinct.)*

Proof. $\Rightarrow$: Suppose that $N \in F[y]$-mod satisfies $\Phi(N) = M$. Replacing N with its elementary-divisor decomposition $\oplus N_i$, we can use the additivity of Φ to assert that $\Phi(N_i)$ is a nonzero submodule of M. But the nonzero submodules of M are precisely the a-scalar modules of size less than or equal to s, so

$$\Phi(N_i) = \{F[x]/(x-a)F[x]\}^{(s_i)},$$

where $\Sigma s_i = s$. Since $N_i = F[y]/p(y)^e F[y]$ for some polynomial $p(y)$, irreducible over F, it follows that $p(y)^e$ divides $y^m - a$. Because $a \neq 0$ and p

does not divide m, $y^m - a$ is separable over F, whence $e = 1$ and $s_i = \dim_F \Psi_x(\Phi(N_i)) = \dim_F \Psi_y(N_i) = \deg p(x)$.

$\Leftarrow$: Let $s = \sum s_i$, where $s_i = \deg p_i(x)$, $p_i(y)$ an irreducible divisor of $y^m - a$. If we define $N = \oplus N_i$, where $N_i = F[y]/p_i(y)F[y]$, it is straightforward to check that $\Phi(N) = M$. This completes the proof. ∎

We can now combine the results of our reductions to piece together the first criterion.

THEOREM 10. *Suppose m is prime to the characteristic of F, and $p(x) \neq x$ is irreducible over F with core $q(x)$ and inseparable degree σ. Let β be a root of $q(x)$, and let E be the simple extension of F by β. Then the $p(x)$-primary module M lies in the image of $\Phi = \Phi_m$ if and only if each of its elementary divisors (the generators of the annihilator ideals of the summands in its elementary-divisor decomposition) occurs with multiplicity which can be expressed as a sum of degrees of irreducible factors over E of the polynomial $x^m - \beta$.*

Proof. M lies in the image of $\Phi \Leftrightarrow$ the $(X^\sigma - \beta)$-primary module $\Omega(M)$ lies in the image of $\Phi: E[Y]\text{-mod} \to E[X]\text{-mod}$ (Theorem 2) $\Leftrightarrow$ the $(x - \beta)$-primary module M' for which $\Gamma_\rho(M') = \Omega(M)$ lies in the image of $\Phi: E[y]\text{-mod} \to E[x]\text{-mod}$ (Theorem 4) $\Leftrightarrow$ each of the homogeneous components of M' lies in the image of Φ (Theorem 7) $\Leftrightarrow$ the multiplicity of each elementary divisor of M' can be expressed as a sum of degrees of irreducible factors over E of the polynomial $x^m - \beta$ (Theorems 8 and 9) $\Leftrightarrow$ the multiplicity of each elementary divisor of M can be expressed as a sum of degrees of irreducible factors over E of the polynomial $x^m - \beta$ (definition of M'). ∎

The matrices $A \in M_n(F)$ for which $M = \text{mod}_x(A)$ form a similarity class which has precisely the same collection of elementary divisors as does M. So, we can say that A is $p(x)$-primary (or homogeneous, or a-scalar) precisely when M is. In matrix language, therefore, Theorem 10 becomes

THEOREM 10′. *Suppose m is prime to the characteristic of F and $p(x) \neq x$ is irreducible over F. Let β be a root of the core of $p(x)$, and let E be the simple extension of F by β. Then the $p(x)$-primary matrix A is an* mth *power in $M_n(F)$ if and only if each of its elementary divisors occurs with multiplicity which can be expressed as a sum of degrees of irreducible factors over E of the polynomial $x^m - \beta$.*

6. THE NILPOTENT CASE

Multiplication by x is a nilpotent endomorphism of the x-primary module M, so we call M a *nilpotent* module. The criteria for deciding when a nilpotent module (nilpotent matrix) lies in the image of Φ [is an mth power in $M_n(F)$] for $F = \mathbf{C}$ [1–3,5] are to a great extent concerned with precisely this case, and practically the same analysis can be used for the general case of arbitrary F.

For any $M \in F[x]$-mod for which there exists an $N \in F[y]$-mod for which $\Phi(N) = M$, the *eigenvalues* of M, i.e., the eigenvalues of any of the matrices A for which $M = \mathrm{mod}_x(A)$, are precisely the mth powers of the eigenvalues of N. Therefore, M is nilpotent if and only if N is. This single observation makes the determination of the criterion in the nilpotent case straightforward.

PROPOSITION 11. *Suppose* $N = F[y]/y^n F[y] \in F[y]$-mod, *so that* $\dim_F N = n$. *Let* q *and* r *be the unique integers that satisfy* $n = qm + r$, $q \geqslant 0$, $0 \leqslant r < m$. *Then* $\Phi(N)$ *is isomorphic to*

$$M = \{F[x]/x^{q+1}F[x]\}^{(r)} \oplus \{F[x]/x^q F[x]\}^{(m-r)}.$$

Proof. The set of vectors $B = \{1, y, y^2, \ldots, y^{n-1}\}$ forms a basis for $V = \Psi_y(N) = \Psi_x(\Phi(N))$ over F (as previously, we write y^i for $y^i + y^n F[y]$). Let B_i, $i = 0, 1, \ldots, m-1$, be the subset of B consisting of the kth powers of y, where $k \equiv i \pmod m$, and let V_i be the subspace of V generated by B_i. Then $xV_i \subset V_i$, since $xy^k = y^{m+k}$, so there is a submodule M_i of $\Phi(N)$ for which $V_i = \Psi_x(M_i)$. Since $\Phi(N) = \sum M_i$ and $M_i \cap M_j = 0$ when $i \neq j$, $\Phi(N) = \oplus M_i$. Further, it is clear that M_i is cyclic in $F[x]$-mod, generated by y^i. But $x^{q+1}y^i = 0$ for all $i < m$, and $x^q y^i = y^{qm+i} \neq 0$ for $i < r$, whence the result. ∎

PROPOSITION 12. *Let* n_1, n_2, q, r_1, r_2 *be integers satisfying* $n_i \geqslant 0$, $q \geqslant 0$, $0 \leqslant r_i < m$, *and* $n_i = qm + r_i$ $(i = 1, 2)$. *Suppose* $r_2 > 0$ *and* $r = \min\{m - r_1, r_2\}$. *Then*

$$\Phi(F[y]/y^{n_1}F[y] \oplus F[y]/y^{n_2}F[y]) = \Phi(F[y]/y^{n_1+r}F[y] \oplus F[y]/y^{n_1-r}F[y]). \tag{8}$$

Proof. By the additivity of Φ and Proposition 11, the $F[x]$-modules in (8) are both equal to

$$\{F[x]/x^{q+1}F[x]\}^{(r_1+r_2)} \oplus \{F[x]/x^qF[x]\}^{(2m-r_1-r_2)}. \quad \blacksquare$$

THEOREM 13. *Suppose the nilpotent module M has elementary-divisor decomposition*

$$M = F[x]/x^{e_1}F[x] \oplus F[x]/x^{e_2}F[x] \oplus \cdots \oplus F[x]/x^{e_t}F[x]$$

and $e_1 \geqslant e_2 \geqslant \cdots \geqslant e_t$. Extend this sequence ad infinitum by putting $e_i = 0$ for all $i \geqslant t$. Then M lies in the image of Φ if and only if for all $i \geqslant 1$,

$$e_{(i-1)m+1} - e_{im} = 0 \text{ or } 1.$$

Proof. Let $M = \Phi(N)$. If

$$N = F[y]/y^{n_1}F[y] \oplus F[y]/y^{n_2}F[y] \oplus \cdots \oplus F[y]/y^{n_s}F[y],$$

$n_1 \geqslant n_2 \geqslant \cdots \geqslant n_s$, then the ith summand will contribute exactly m summands to the elementary-divisor decomposition of M unless $n_i < m$, in which case it contributes exactly n_i summands of the form $F[x]/xF[x]$. By appending to these n_i summands $m - n_i$ trivial summands $F[x]/x^0F[x] = 0$, we may write for each $i \geqslant 1$

$$\Phi(F[y]/y^{n_i}F[y]) = F[x]/x^{d_{i1}}F[x] \oplus F[x]/x^{d_{i2}}F[x] \oplus \cdots \oplus F[x]/x^{d_{im}}F[x],$$

or $\Phi(n_i) = d_{i1} \oplus d_{i2} \oplus \cdots \oplus d_{im}$ for short. Here, we extend the sequence $n_1 \geqslant n_2 \geqslant \cdots \geqslant n_s$ by setting $n_i = 0$ for $i > s$. Also, we are free to assume that $d_{i1} \geqslant d_{i2} \geqslant \cdots \geqslant d_{im}$ for all i. The sequence

$$d_{11}, d_{12}, \ldots, d_{1m}, d_{21}, \ldots \tag{9}$$

is precisely the sequence $e_1, e_2, \ldots$ in some order. Further, it follows from Proposition 11 that for all i, $d_{i1} - d_{im} = 0$ or 1; in fact, $n_i = d_{im}m + r_i$, where r_i is the multiplicity of d_{i1} in the subsequence $d_{i1}, d_{i2}, \ldots, d_{im}$. Therefore, each of the first s subsequences $d_{i1}, d_{i2}, \ldots, d_{im}$ of (9) is nonincreasing. So we are done if we can show that the n_i can be chosen so that for

all $i < s$, $d_{im} \geqslant d_{i+1,1}$ for then (9) is nonincreasing, hence is exactly equal to $e_1, e_2, \ldots$.

Now if for some i, $d_{im} < d_{i+1,1}$, then, putting $q = d_{im}$, we must have $n_i = qm + r_i$ and $n_{i+1} = qm + r_{i+1}$ and can apply Proposition 12. Specifically, where $r = \min\{m - r_1, r_2\} > 0$, $\Phi(n_i) \circledast \Phi(n_{i+1}) = \Phi(n_i + r) \oplus \Phi(n_{i+1} - r)$, and either $\Phi(n_i + r) = (q+1) \oplus (q+1) \oplus \cdots \oplus (q+1)$ and $\Phi(n_{i+1} - r) = (q+1) \oplus \cdots \oplus (q+1) \oplus q \oplus \cdots \oplus q$ ($q+1$ occurring $r - r_2$ times) when $r = m - r_1$, or $\Phi(n_i + r) = (q+1) \oplus \cdots \oplus (q+1) \oplus q \oplus \cdots \oplus q$ (q occurring $r - [m - r_1]$ times) and $\Phi(n_{i+1} - r) = q \oplus q \oplus \cdots \oplus q$ when $r = r_2$. In either case, we see that replacing $n_i \geqslant n_{i+1}$ with $n_i + r \geqslant n_{i+1} - r$ ensures that $d_{im} \geqslant d_{i+1,1}$, but may violate $n_1 \geqslant n_2 \geqslant \cdots$. However, it should be clear that after finitely many applications of this replacement procedure, we do obtain $n_1 \geqslant n_2 \geqslant \cdots$ for which the associated sequence (9) is nonincreasing, completing the proof. ∎

In terms of matrices, this becomes

THEOREM 13′. *The nilpotent matrix $A \in M_n(F)$ has elementary divisors which are all powers of x; let these exponents of x be $e_1 \geqslant e_2 \geqslant \cdots \geqslant e_t$, where e_t is the last nonzero exponent. Make this an infinite sequence by defining $e_i = 0$ for all $i \geqslant t$. Then A is an mth power in $M_n(F)$ if and only if for all $i = 1, 2, \ldots,$*

$$e_{(i-1)m+1} - e_{im} = 0 \text{ or } 1.$$

7. p-POWER ROOTS IN CHARACTERISTIC p

The only remaining case we need to consider is the case of the $(x-a)$-primary module, $a \neq 0$, when m is divisible by p. Indeed, if $m = n\rho$, $\rho = p^r$, $r \geqslant 1$, n prime to p, then we may use the fact that $\Phi_m = \Phi_n \Phi_\rho$ to reduce to the case $n = 1$, and then handle the more general situation with the results of Section 5.

THEOREM 15. *Suppose the module M has elementary-divisor decomposition*

$$M = F[x]/(x-a)^{e_1}F[x] \oplus F[x]/(x-a)^{e_2}F[x]$$

$$\oplus \cdots \oplus F[x]/(x-a)^{e_t}F[x]$$

and that $e_1 \geqslant e_2 \geqslant \cdots \geqslant e_t$. *Extend this sequence ad infinitum by putting* $e_j = 0$ *for all* $j > t$. *Let* $\tau \leqslant \rho$ *be the largest power of* p *for which* a *is a* τth *power in* F, *and put* $\sigma = \rho/\tau$. *Then* M *lies in the image of* Φ_ρ: $F[y]$-mod $\to$ $F[x]$-mod *if and only if the multiplicity of each* e_j *is a multiple of* σ *and for all* $i \geqslant 1$,

$$e_{(i-1)\rho+1} - e_{i\rho} = 0 \text{ or } 1.$$

Proof. $\Rightarrow$: Let $M = \Phi_\rho(N)$, and suppose that $N = F[y]/p(y)^e F[y]$, $p(y)$ an irreducible over F. As an element of $\mathrm{End}_F V$, $V = \Psi_x(M) = \Psi_y(N)$, $(y^\rho - a)^{e_1} = (x-a)^{e_1}$ is the zero map, so the polynomial $(y^\rho - a)^{e_1}$ must be divisible by the minimal polynomial of y, $p(y)^e$. But $(y^\rho - a)^{e_1}$ has exactly one root α, of multiplicity $e_1\rho$, so it factors over F as $(y^{\rho/\tau} - b)^{\tau e_1}$, where $a = b^\tau$. If $p(y)$ has separable core $q(y)$, then it follows that $q(y) = y - b$ and that $\sigma = \rho/\tau$ is its inseparable degree. So $N = F[y]/(y^{\rho/\tau} - b)^e F[y]$. Therefore,

$$\begin{aligned} M = \Phi_\rho(N) &= \Phi_\tau\Phi_\sigma(N) \\ &= \Phi_\tau\big(F[z]/q(z)^e F[z]\big)^{(\sigma)} \\ &= \Phi_\tau\big(F[z]/(z-b)^e F[z]\big)^{(\sigma)} \end{aligned}$$

by Proposition 3, viewing Φ_ρ as the composition

$$F[y]\text{-mod} \xrightarrow{\Phi_\sigma} F[z]\text{-mod} \xrightarrow{\Phi_\tau} F[x]\text{-mod}.$$

Put $W = \Psi_x(\Phi_\tau(F[z]/(z-b)^e F[z])) = \Psi_z(F[z]/(z-b)^e F[z])$; it has basis $B = \{1, z-b, (z-b)^2, \ldots, (z-b)^{e-1}\}$ (writing $(z-b)^i$ for $(z-b)^i + (z-b)^e F[z]$). Multiplication by $x - a = z^\tau - a = (z-b)^\tau$ then leaves invariant the subspaces W_i with respective bases $B_i = \{(z-b)^j \mid j \equiv i \pmod\tau\}$, $i = 0, 1, \ldots, \tau - 1$. So there are submodules L_i of $\Phi_\tau(F[z]/(z-b)^e F[z])$, cyclic with generators $(z-b)^i$, for which $W_i = \Psi_x(L_i)$. As $W = \oplus W_i$, we have $\Phi_\tau(F[z]/(z-b)^e F[z]) = \oplus L_i$. Let c, d be the integers satisfying $e = c\tau + d$, $c \geqslant 0$, $0 \leqslant d < \tau$. Since $(x-a)^{c+1}(z-b)^i = 0$ for all $i < \tau$, but $(x-a)^c(z-b)^i = 0$ only for $i < d$, it follows that $L_i = F[x]/(x-a)^{c+1}F[x]$ for $i < d$ and $L_i = F[x]/(x-a)^c F[x]$ for $i \geqslant d$.

Therefore,

$$\begin{aligned} M &= \Phi_\tau\big(F[z]/(z-b)^e F[z]\big)^{(\sigma)} \\ &= \big\{F[x]/(x-a)^{c+1}F[x]\big\}^{(d\sigma)} \oplus \big\{F[x]/(x-a)^c F[x]\big\}^{(\rho - d\sigma)}. \end{aligned} \tag{10}$$

From (10), the multiplicity of each summand of M is a multiple of σ and the criteria of the theorem are satisfied. In general, if N has s summands in its

elementary-divisor decomposition, we can use (10) and the additivity of Φ_ρ to complete this part of the proof.

$\Leftarrow$: Suppose the multiplicity of each e_j is a multiple of σ and for all $i \geqslant 1$, $e_{(i-1)\rho+1} - e_{i\rho} = 0$ or 1. Then the sequence of integers $d_1 \geqslant d_2 \geqslant \cdots$ defined by

$$d_i = \sigma^{-1} \sum_{k=1}^{\rho} e_{(i-1)\rho+k}$$

is such that

$$N = \oplus F[y]/(y^\sigma - b)^{d_i} F[y]$$

satisfies $\Phi_\rho(N) = M$: by Proposition 3 and (10),

$$\begin{aligned}
\Phi_\rho(N) &= \oplus \Phi_\rho\big(F[y]/(y^\sigma - b)^{d_i} F[y]\big) \\
&= \oplus \Phi_\tau \Phi_\sigma\big(F[y]/(y^\sigma - b)^{d_i} F[y]\big) \\
&= \oplus \Phi_\tau\big(F[z]/(z - b)^{d_i} F[z]\big)^{(\sigma)} \\
&= \oplus \big\{F[x]/(x-a)^{e_{(i-1)\rho+1}} F[x] \oplus F[x]/(x-a)^{e_{(i-1)\rho+2}} F[x] \\
&\qquad\qquad \oplus \cdots \oplus F[x]/(x-a)^{e_{i\rho}} F[x]\big\} \\
&= M. \qquad \blacksquare
\end{aligned}$$

Observe that the similarity of Theorem 15 (and its proof) to Theorem 13 (and its proof) is a result of the fact that the ρth roots of a are all identical, just as the mth roots of 0 (m arbitrary) are all identical; the opposite holds in the situation of Theorem 8: when m is prime to p, the mth roots of a are all distinct. This is not to say that there is no connection between the criteria of Theorem 8 and 15. Indeed, the criterion that determines when the general $p(x)$-primary module is in the image of Φ_ρ can be put into a form that incorporates both.

THEOREM 16. *Suppose that $\rho = p^r$ and that $p(x) \neq x$ is an irreducible polynomial over F with core $q(x)$ and inseparable degree $\sigma = p^s$, that β is a root of $q(x)$, and that E is the simple extension field $F(\beta)$. Then the*

$p(x)$-primary module M as in (1) *lies in the image of Φ_ρ if and only if each of its elementary divisors occurs with multiplicity which can be expressed as a sum of degrees of irreducible factors over E of the polynomial $x^\rho - \beta$, and for all $i \geqslant 1$,*

$$e_{(i-1)\rho+1} - e_{i\rho} = 0 \text{ or } 1,$$

where $e_1 \geqslant e_2 \geqslant \cdots$ is the sequence of exponents in (1) *extended by putting $e_j = 0$ for all $j > t$.*

Proof. Let $\tau \leqslant \rho$ be the largest power of p for which $\beta = \gamma^\tau$, $\gamma \in E$; put $\sigma = \rho/\tau$. Then M lies in the image of $\Phi_\rho \Leftrightarrow$ the $(X^\sigma - \beta)$-primary module $\Omega(M)$ lies in the image of $\Phi_\rho: E[Y]\text{-mod} \to E[X]\text{-mod}$ (Theorem 2) $\Leftrightarrow$ the $(x - \beta)$-primary module M' for which $\Gamma_\rho(M') = \Omega(M)$ lies in the image of $\Phi_\rho: E[y]\text{-mod} \to E[x]\text{-mod}$ (Theorem 4) $\Leftrightarrow$ each e_j among the sequence of exponents of the elementary divisors of M', and hence of M (both M and M' have the same sequence of exponents), has multiplicity a multiple of σ, and for all $i \geqslant 1$, $e_{(i-1)\rho+1} - e_{i\rho} = 0$ or 1 (Theorem 15) $\Leftrightarrow$ each e_j among the sequence of exponents of the elementary divisors of M has multiplicity which can be expressed as a sum of degrees of irreducible factors over E of the polynomial $x^\rho - \beta$, and for all $i \geqslant 1$, $e_{(i-1)\rho+1} - e_{i\rho} = 0$ or 1 [as $x^\rho - \beta$ factors into irreducibles over E as $(x^\sigma - \gamma)^\tau$]. ∎

In terms of matrices:

THEOREM 16′. *Suppose that $\rho = p^r$ and that $p(x) \neq x$ is an irreducible polynomial over F with core $q(x)$ and inseparable degree $\sigma = p^s$, that β is a root of $q(x)$, and that E is the simple extension field $F(\beta)$. Then the $p(x)$-primary matrix $A \in M_n(F)$ is a ρth power in $M_n(F)$ if and only if each of its elementary divisors occurs with multiplicity which can be expressed as a sum of degrees of irreducible factors over E of the polynomial $x^\rho - \beta$, and for all $i \geqslant 1$,*

$$e_{(i-1)\rho+1} - e_{i\rho} = 0 \text{ or } 1,$$

where $e_1 \geqslant e_2 \geqslant \cdots$ is the sequence of exponents in (1) *extended by putting $e_j = 0$ for all $j > t$.*

To conclude our analysis, we ought to apply Theorems 10 and 16 to the general case of describing the image of $\Phi_m = \Phi_n\Phi_\rho$ when $m = n\rho$, n prime to p. However, this information is not enough. To determine whether M lies in the image of $\Phi_n\Phi_\rho$, we need to be able to identify whether the preimage L of

M under Φ_ρ lies in the image of Φ_n, and this requires knowing the structure of all possible preimage modules L. This seems to be a difficult problem, in general. (Alternatively, we need to be able to identify whether the preimage of a module M under Φ_n lies in the image of Φ_ρ, but this is even more difficult; the behavior of Φ_ρ is much simpler than that of Φ_n.)

8. MORE QUESTIONS

Some obvious questions arise:

1. Is there an algorithm which will find all the mth roots of A, so long as at least one exists? In module-theoretic form, this is precisely the problem we had in Section 7 in identifying all the preimages under Φ_ρ of a given module M which we know to lie in the image.

2. If this algorithm exists, for which matrices A is it effective? For instance, one can actually exhibit up to similarity all the mth roots of any nilpotent matrix that satisfies the criterion of Theorem 13′. In principle, it is also possible to exhibit the mth roots of arbitrary matrices with complex entries (see the examples in [3]), and numerical analysts run a small industry in calculating roots of real matrices (see for instance [11] and [12]). This may not be feasible, however, in the case of $p(x)$-primary matrices with rational entries for which $\deg p(x)$ is large [so that factoring $x^m - \beta$ over $\mathbf{Q}(\beta)$ is difficult].

3. Hodges [6] counts the solutions to $X^2 = I$ over finite fields. In general, can we find the number of (similarity classes) of solutions to $X^m = A$, i.e., the number of (isomorphism classes) of modules N that satisfy $\Phi_m(N) = M$? This is probably not as difficult as finding all the solutions, being a problem of a more combinatorial than algebraic nature. What connections can be made with the fundamental theorem of algebra; that is, in what sense does the number of solutions depend on m?

4. An interesting side question: in Section 3, we proved that the map $\omega: F[x]/p(x)^e F[x] \to E[X]/(X^\rho - \beta)^e E[X]$ which sends $f(x) + p(x)^e F[x]$ to $f(X) + (X^\rho - \beta)^e E[X]$ is a ring isomorphism. What explicitly is the preimage of β?

The author must acknowledge the advice and direction that Professor David Sibley, Professor Don James, and Carsten Hansen offered during the course of this investigation (which in preliminary form served as his Ph.D. thesis [16]). Special thanks must also go to Professor Mark Kleiner, whose suggestions about the style helped me to present these results in a more modern and elegant form.

REFERENCES

1 J. H. M. Wedderburn, *Lectures on Matrices*, Dover, New York, 1964.

2 F. R. Gantmacher, *Matrix Theory*, Vol. 1, transl. by K. A. Hirsch, Chelsea, New York, 1960.

3 Edward T. Browne, *Introduction to the Theory of Determinants and Matrices*, Univ. of North Carolina Press, Chapel Hill, 1958.

4 Harvey Flanders, Analytic solutions of matrix equations, *Linear and Multilinear Algebra* 2:241–243 (1974).

5 G. W. Cross and P. Lancaster, Square roots of complex matrices, *Linear and Multilinear Algebra* 1:289–293 (1974).

6 John H. Hodges, The matrix equation $X^2 - I = 0$ over a finite field, *Amer. Math. Monthly* 65:518–520 (1958).

7 G. Alefeld and N. Schneider, On square roots of M-matrices, *Linear Algebra Appl.* 42:119–132 (1982).

8 S. K. Jain, V. K. Goel, and Edward K. Kwak, Nonnegative mth roots of nonnegative 0-symmetric idempotent matrices, *Linear Algebra Appl.* 23:37–51 (1979).

9 Nathan Jacobson, *Basic Algebra I*, 3rd ed., Freeman, New York, 1974.

10 Serge Lang, *Algebra*, rev., Addison-Wesley, Reading, Mass., 1971.

11 Eugene D. Denman, Roots of real matrices, *Linear Algebra Appl.* 36:133–139 (1981).

12 Åke Björck and Sven Hammarling, A Schur method for the square root of a matrix, *Linear Algebra Appl.* 52/53:127–140 (1983).

13 R. P. Bambah and S. Chowla, On integer cube roots of the unit matrix, *Sci. and Culture* 12:69–70 (1946).

14 Saunders MacLane, *Categories for the Working Mathematician*, Graduate Texts in Math. 5, Springer-Verlag, New York, 1971.

15 D. W. Masser and M. Neumann, On the square roots of strictly quasiaccretive complex matrices, *Linear Algebra Appl.* 28:135–140 (1979).

16 Daniel E. Otero, Extraction of mth Roots in Matrix Rings over Fields, Ph.D. Thesis, Pennsylvania State Univ., University Park, 1987.

17 H. J. Ryser, A generalization of the matrix equation $A^2 = J$, *Linear Algebra Appl.* 3:451–460 (1970).

Received 17 June 1988; final manuscript accepted 11 November 1988